

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-132480

(P2000-132480A)

(43) 公開日 平成12年5月12日 (2000.5.12)

(51) Int.Cl.<sup>7</sup>

G 0 6 F 13/00

識別記号

3 5 4

F I

G 0 6 F 13/00

テマコード\* (参考)

3 5 4 D 5 B 0 8 9

審査請求 未請求 請求項の数 8 O L (全 7 頁)

(21) 出願番号 特願平10-305473

(22) 出願日 平成10年10月27日 (1998.10.27)

(71) 出願人 000004226

日本電信電話株式会社

東京都千代田区大手町二丁目3番1号

(72) 発明者 渡辺 一成

東京都新宿区西新宿三丁目19番2号 日本  
電信電話株式会社内

(74) 代理人 100100893

弁理士 渡辺 勝 (外1名)

Fターム(参考) 5B089 GA21 HA10 JA22 JB06 KA12

KB06 KC14 KC47 KH02 LB14

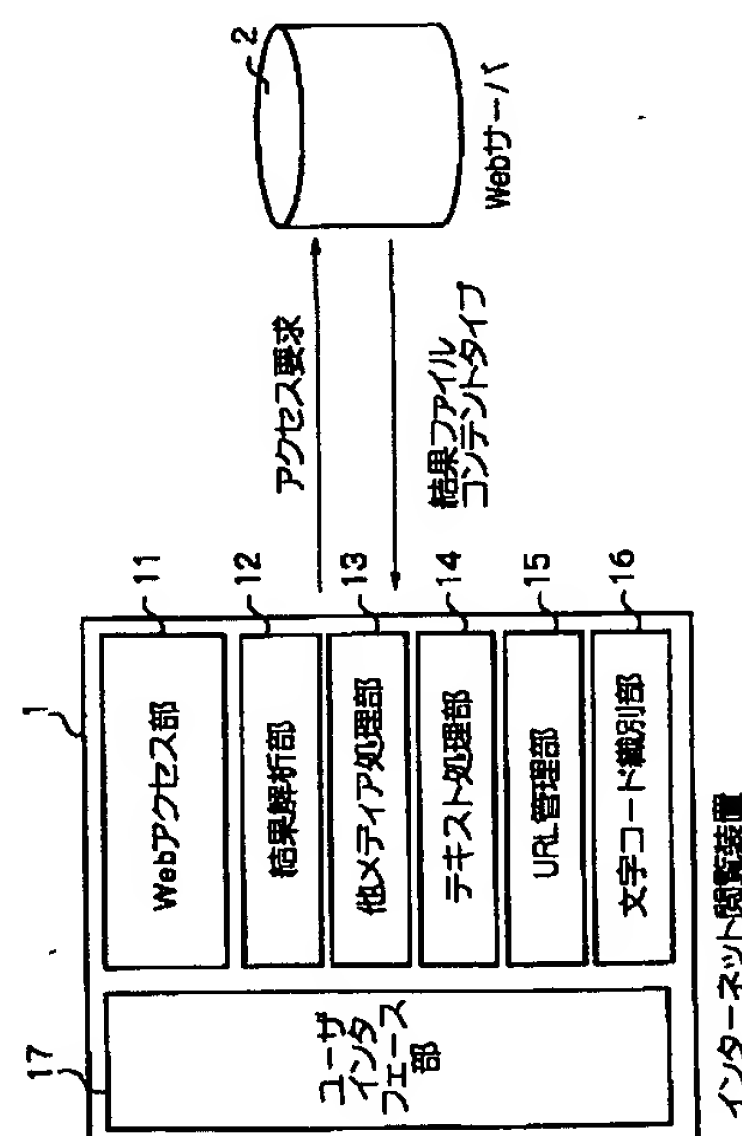
MC11

(54) 【発明の名称】 インターネット閲覧方法、装置、およびインターネット閲覧プログラムを記録した記録媒体

(57) 【要約】

【課題】 インターネット閲覧において、設定を変更することなく文字化けのない表示を行う。

【解決手段】 利用者からWebサーバ2へのアクセス要求がWebサーバ11からWebサーバ2へ送られると、結果ファイルとコンテンツタイプがWebサーバ2から返却され、Webアクセス部11で受信される。結果解析部12の解析により、返却された結果ファイルがテキストの場合、テキスト処理部14は当該ファイル中に文字コード、文字セットが明示的に記述されているかどうかチェックする。明示的に記述されていない場合、URL管理部15と文字コード識別部16により文字コードと文字セットの識別処理が行われ、識別結果に基づいてテキストファイルがユーザインタフェース部17に表示される。



利用者

## 【特許請求の範囲】

【請求項 1】 Webサーバにアクセス要求を発行し、その結果として前記Webサーバから返却されるファイルを受信し、表示するインターネット閲覧方法において、受信したファイル中に文字コード、文字セットが明示的に記述されているかどうか解析し、明示的に記述されていない場合には文字コード、文字セットを識別した後に前記ファイルの表示を行うことを特徴とするインターネット閲覧方法。

【請求項 2】 利用者が指定したWebサーバへアクセスし、これに回答して前記Webサーバが返却した結果ファイルとコンテンツタイプ情報を受信するWebアクセス段階と、

前記コンテンツタイプ情報を判定する結果解析段階と、前記コンテンツタイプ情報がテキストの場合、前記結果ファイル中に文字コード、文字セットが明示的に記述されているかどうかのチェックを行い、明示的に記述されているならば、前記文字コード、前記文字セットの情報とともに当該テキストファイルを出力するテキスト処理段階と、

前記結果ファイル中に文字コード、文字情報が明示的に示されていない場合、前記結果ファイルを検査してその文字コード、文字セットを識別し、識別結果と当該テキストファイルを出力する文字コード識別段階と、

前記コンテンツタイプ情報がテキスト以外の場合、該コンテンツタイプ情報に対応する処理を行う他メディア処理段階と、

出力されたテキストファイルを文字コード、文字セットの情報または識別結果に基づいて表示する表示段階を有するインターネット閲覧方法。

【請求項 3】 前記文字コード識別段階は、識別結果と当該テキストファイルのURL情報を識別済みURL DBに記録する段階を有し、前記テキスト処理段階から出力された識別対象のテキストファイルのURLと、前記識別済みURL DBに格納されているURLの比較を行い、前者のURLのディレクトリ階層が後者のURLより下位階層である場合、前記識別済みURL DB中の識別結果と前記テキストファイルを前記表示段階に渡し、それ以外の場合には前記文字コード識別段階による文字コード、文字セットの識別を行うURL管理段階をさらに有する請求項 2 記載の方法。

【請求項 4】 Webサーバにアクセス要求を発行し、その結果としてWebサーバから返却されるファイルを受信し、表示するインターネット閲覧装置において、受信したファイル中に文字コード、文字セットが明示的に記述されているかどうか解析する手段と、明示的に記述されていない場合には文字コード、文字セットを識別する手段と、識別結果に基づいて前記ファイルの表示を行う手段を有することを特徴とするインターネット閲覧装置。

【請求項 5】 利用者が指定したWebサーバへアクセスし、これに回答して前記Webサーバが返却した結果ファイルとコンテンツタイプ情報を受信するWebアクセス手段と、

前記コンテンツタイプ情報を判定する結果解析手段と、前記コンテンツタイプ情報がテキストの場合、前記結果ファイル中に文字コード、文字セットが明示的に記述されているかどうかのチェックを行い、明示的に記述されているならば、前記文字コード、前記文字セットの情報とともに当該テキストファイルを出力するテキスト処理手段と、

前記結果ファイル中に文字コード、文字情報が明示的に示されていない場合、前記結果ファイルを検査してその文字コード、文字セットを識別し、識別結果と当該テキストファイルを出力する文字コード識別手段と、

前記コンテンツタイプ情報がテキスト以外の場合、該コンテンツタイプ情報に対応する処理を行う他メディア処理手段と、

前記Webサーバへの利用者のアクセス要求を受け、また前記の出力されたテキストファイルを文字コード、文字セットの情報または識別結果に基づいて表示するユーザインタフェース手段を有するインターネット閲覧装置。

【請求項 6】 識別済みURL DBをさらに有し、前記文字コード識別手段は、識別結果と当該テキストファイルのURL情報を前記識別済みURL DBに記録する手段を有し、前記テキスト処理手段から出力された識別対象のテキストファイルのURLと、前記識別済みURL DBに格納されているURLの比較を行い、前者のURLのディレクトリ階層が後者のURLより下位階層である場合、前記識別済みURL DB中の識別結果と前記テキストファイルを前記ユーザインタフェース手段に渡し、それ以外の場合には前記文字コード識別手段による文字コード、文字セットの識別を行うURL管理手段をさらに有する請求項 5 記載の装置。

【請求項 7】 利用者が指定したWebサーバへアクセスし、これに回答して前記Webサーバが返却した結果ファイルとコンテンツタイプ情報を受信するWebアクセス処理と、

前記コンテンツタイプ情報を判定する結果解析処理と、前記コンテンツタイプ情報がテキストの場合、前記結果ファイル中に文字コード、文字セットが明示的に記述されているかどうかのチェックを行い、明示的に記述されているならば、前記文字コード、前記文字セットの情報とともに当該テキストファイルを出力するテキスト処理と、

前記結果ファイル中に文字コード、文字情報が明示的に示されていない場合、前記結果ファイルを検査してその文字コード、文字セットを識別し、識別結果と当該テキストファイルを出力する文字コード識別処理と、

前記コンテンツタイプ情報がテキスト以外の場合、該コンテンツタイプ情報に対応する処理を行う他メディア処理と、

前記テキスト処理または文字コード識別処理から出力されたテキストファイルを文字コード、文字セットの情報または識別結果に基づいて表示装置に表示する表示処理をコンピュータに実行させるためインターネット閲覧プログラムを記録した記録媒体。

【請求項8】 前記文字コード識別処理は、識別結果と当該テキストファイルのURL情報を識別済みURLDBに記録する処理を有し、前記テキスト処理から出力された識別対象のテキストファイルのURLと、前記識別済みURLDBに格納されているURLの比較を行い、前者のURLのディレクトリ階層が後者のURLより下位階層である場合、前記識別済みURLDB中の識別結果と前記テキストファイルを前記表示処理に渡し、それ以外の場合には前記文字コード識別処理による文字コード、文字セットの識別を行うURL管理処理をさらに有する請求項7記載の記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、Webサーバにアクセス要求を発行し、その結果としてWebサーバから返却されるファイルを受信し、表示するインターネット閲覧方法および装置に関する。

【0002】

【従来の技術】従来のインターネット閲覧装置ではWebサーバからファイルを受信すると、まずファイル中に文字コード、文字セットが明示的に指定されているかどうか調べる。指定されている場合はその指定に従ってファイルの表示を行う。指定されていない場合は利用者が閲覧装置で予め設定した文字コード、文字セット情報を利用しそれに従った表示を行う。この場合、利用者設定と実際に受信したHTMLファイルの文字コード、文字セットが異なる場合には文字化けとなるが、利用者がその都度設定を変更することで対処する。

【0003】Webサーバから返却される全てのファイルについてHTML (Hyper Text Markup Language)等の言語により文字コード、文字セットが明示的に指定されていれば閲覧装置側はその指定に従い表示することで文字化けは回避できる。上記問題が発生するのは実際のファイルではそのような指定は省略されており、閲覧装置側で表示するための情報が得られないためである。例えば、HTMLで規定されているMETAタグはその属性としてcharset (character set)属性を持つことができ、その属性値で文字コードと文字セットを指定できる。これらはそのHTMLファイルがどのような文字コード、文字セットで記述してあるかを明示的に示すためのものである。閲覧装置ではMETA

タグとcharset属性があればその文字コード、文字セットに従った表示を行うことで利用者の設定に頼らず文字化けのない表示ができる。しかし、全てのHTMLファイルにMETAタグ、charset属性が指定してあれば文字化けの問題は起らないが、現状のWebサーバ中のHTMLファイルにおいてこのMETAタグとcharset属性を指定してあるのはごく少数であり、そのために閲覧装置側では文字コード、文字セットを判断できず利用者の設定による表示を行っている。

【0004】

10 【発明が解決しようとする課題】上述した従来のインターネット閲覧装置では、ファイル中に文字コード、文字セットの明示的な指定がない場合に利用者設定と実際のファイルのそれが異なると、文字化けが発生する。このような場合、利用者はその設定を試行錯誤的に変更し、文字化けが起らないような設定をする必要がある。しかしながら、多言語環境でのインターネット閲覧では利用者がその設定を変更する状況が頻繁に発生するため設定負荷が問題となる。そこで設定負荷を極力減らすような機構が必要である。

20 【0005】本発明の目的は、設定を変更することなく文字化けのない表示が可能なインターネット閲覧方法、装置、およびインターネット閲覧プログラムを記録した記録媒体を提供することにある。

【0006】

【課題を解決するための手段】本発明は、受信したファイル中に文字コード、文字セットが明示的に記述されているかどうか解析し、記述されていない場合には文字コード、文字セットを識別した後にファイルの表示を行うものである。

30 【0007】なお、この方法は識別のためのオーバーヘッドがあり、従来の方法に比べて閲覧処理速度が遅くなる欠点がある。そこで、参照しているURL (Uniform Resource Locator)中のディレクトリ階層を利用し、直前に上位のディレクトリ階層で文字コード、文字セットの識別が行われている場合にはそれより下位のディレクトリ階層の参照においては上位の識別結果を利用することで処理の高速化をはかる。

40 【0008】このように、本発明では、Webサーバから返却されるファイルに明示されている情報、またはファイル自体を走査して得られる文字コード、文字セットの情報を利用するため、利用者の設定に頼ることなく文字化けのない表示を行うことが可能になる。結果として多言語環境においても利用者の設定負荷を大幅に削減することができる。万が一、識別処理に失敗し文字化けしても、従来の方法により最低限の設定負荷で表示が可能になる。

【0009】

【発明の実施の形態】次に、本発明の実施の形態について図面を参照して説明する。

50 【0010】図1を参照すると、本発明の一実施形態の

インターネット閲覧システムはインターネット閲覧装置 1 と Web サーバ 2 で構成されている。ここで、Web サーバ 2 から返却されるファイルは HTML ファイルとする。

【0011】閲覧装置 1 は Web アクセス部 11 と結果解析部 12 と他メディア処理部 13 とテキスト処理部 14 と URL 管理部 15 と文字コード識別部 16 とユーザインタフェース部 17 で構成されている。

【0012】利用者が閲覧装置で目的とする Web サーバ 2 の URL をユーザインタフェース部 17 を介して指定すると、Web アクセス部 11 が Web サーバ 2 に対してアクセスを行う。Web サーバ 2 からは結果ファイルと各ファイルに対応するコンテンツタイプ情報が返却され、Web アクセス部 11 が受信する。これらは結果解析部 12 に送られる。図 2 に Web サーバ 2 から返却されるファイルと対応するコンテンツタイプの例を示す。

【0013】図 3 は結果解析部 12 の処理を示す図である。結果解析部 12 では Web サーバ 2 から返却されたコンテンツタイプ情報を判定し（ステップ 31～35）、ファイルをテキスト処理部 14 または他メディア処理部 13 の対応する処理部 21～24 に割り振る。図 2 のファイルを例にとると、result.html はコンテンツタイプ情報が text/html であるのでテキスト処理部 14 へ、image1.gif、image2.gif はコンテンツタイプ情報が image/gif であるのでイメージ処理部 21 に各々渡される。

【0014】図 4 にテキスト処理部 14 の処理を示す。テキスト処理部 14 ではまず渡されたファイル中で文字コード、文字セットが明示的に記述されているかのチェックを行う。例えば HTML ファイルの場合には META タグの charset 属性の指定の有無をチェックする（ステップ 41、42）。図 5 に HTML ファイルの META タグの例を示す。指定されている場合にはそこから得られる文字コード、文字セットの情報とともに当該テキストファイルをユーザインタフェース部 17 に渡す。ユーザインタフェース部 17 ではテキストファイルを渡された文字コード、文字セットの情報に基づいて表示することで文字化けのない表示を行う。指定されていない場合には URL 管理部 15 と文字コード識別部 16 により識別処理が行われる。識別結果はユーザインタフェース部 17 に送られる。

【0015】図 6 に URL 管理部 15 と文字コード識別部 16 の処理を示す。URL 管理部 15 では渡されたファイルの URL のチェックを行う。識別済み URL DB 18 中には直前に文字コード識別部 16 で識別された URL と識別結果が格納されている。URL 管理部 15 は識別済み URL DB 18 にアクセスし、識別対象となっているファイルの URL と DB 18 に格納されている URL の比較を行う。識別対象ファイルの URL ディ

レクトリ階層が識別済み URL DB 18 中の URL より下位階層である場合には同一の文字コード、文字セットを使用している可能性が高いと考えられる。この場合、識別済み URL DB 18 中の識別結果（文字コード、文字セット情報）と識別対象ファイルをユーザインタフェース部 17 に渡し表示処理を行う。それ以外の場合（識別済み URL DB 18 が空の場合も含む）は文字コード識別部 16 による識別を行う。図 7 に URL 管理部 15 と識別済み URL DB 18 の処理例を示す。

【0016】文字コード識別部 16 では渡されたファイルを検査してその文字コード、文字セットを識別する。この処理には例えば特開平 9-212334 で記述されている方法が利用できる。識別結果と当該ファイルはユーザインタフェース部 17 に渡され表示処理が行われる。また、同時に識別結果と当該ファイルの URL 情報は識別済み URL DB 18 に記録される。このように最終的にユーザインタフェース部 17 に表示するファイルとともに文字コード、文字セット情報をユーザインタフェース部 17 に与えることで、利用者の設定によらず文字化けのない表示が可能となる。

【0017】図 8 は図 1 に示したインターネット閲覧装置 1 の、記録媒体を用いた実施形態を示すもので、入力装置 51 と記憶装置 52、53 と表示装置 54 と送信装置 55 と記録媒体 56 とデータ処理装置 57 で構成されている。

【0018】入力装置 51 は利用者がアクセス要求を入力するためのものである。記憶装置 52 はハードディスクである。記憶装置 53 は識別済み URL DB 18 に相当する。表示装置 54 は Web サーバ 2 から返却された結果ファイルを表示する。送信装置 55 はアクセス要求を Web サーバ 2 に送信するためのものである。記憶媒体 56 はインターネット閲覧装置 1 の各部 11～16 の処理からなるインターネット閲覧プログラムが記録されている、フロッピー・ディスク、CD-ROM、光磁気ディスク、半導体メモリ等の記録媒体である。データ処理装置 57 は記録媒体 56 からインターネット閲覧プログラムを記憶装置 52 に読み込んで、これを実行する CPU である。

【0019】

【発明の効果】以上説明したように、本発明によれば、利用者が当インターネット閲覧装置の設定を変更することなく文字化けのない表示が可能であり、特に多言語環境においては設定のための負荷を大幅に低減することができる。

【図面の簡単な説明】

【図 1】本発明の一実施形態のシステム構成図である。

【図 2】Web サーバ 2 から返却されるファイルの例を示す図である。

【図 3】図 1 中の結果解析部 12 の処理を示す図である。

【図4】図1中のテキスト処理部14の処理を示す図である。

【図5】HTMLファイル中のMETAタグの例を示す図である。

【図6】URL管理部15と文字コード識別部16の処理を示す図である。

【図7】URL管理部15と識別済みURL DB18の処理例を示す図である。

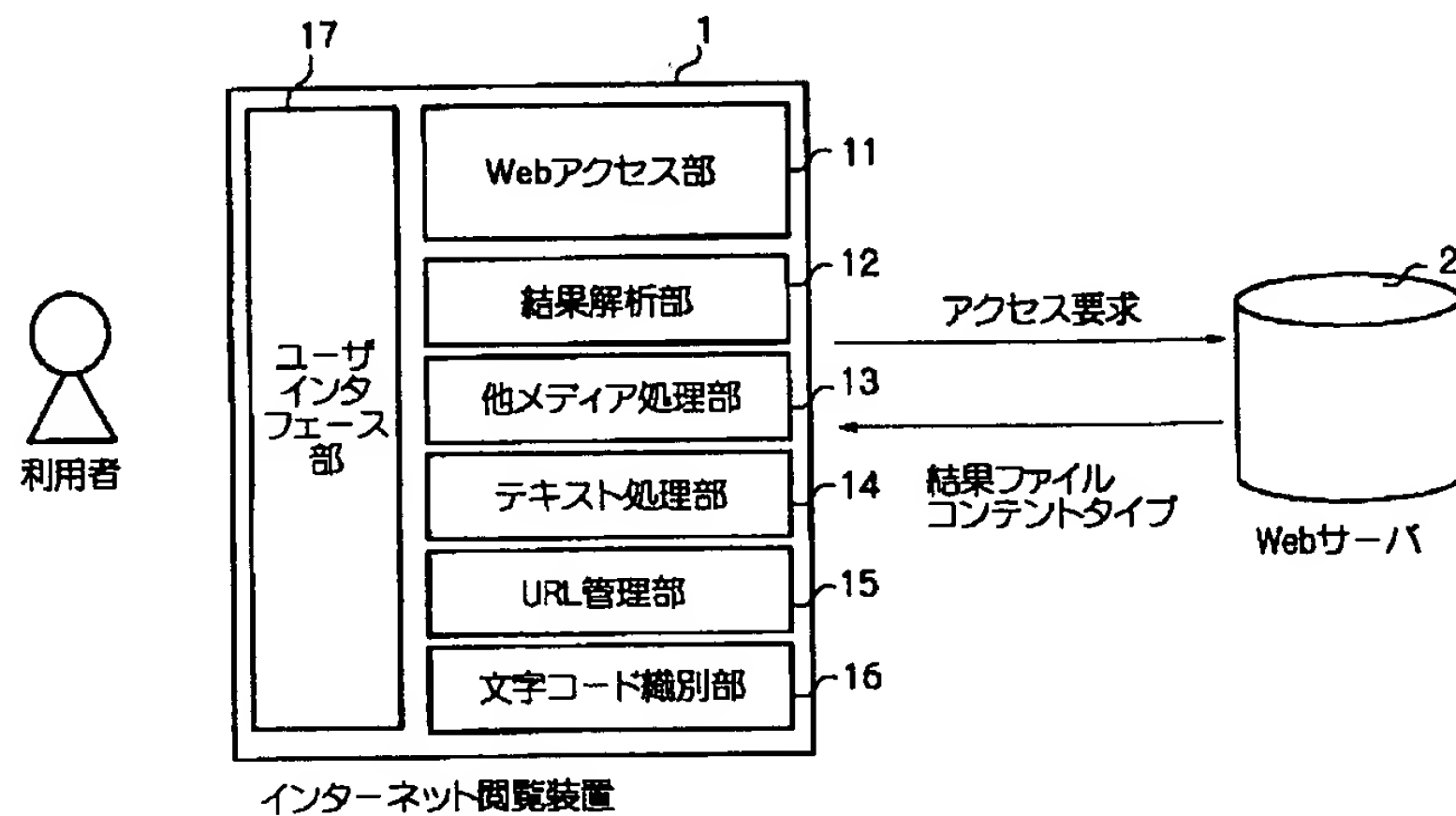
【図8】本発明の他の実施形態のインターネット閲覧装置の構成図である。

【符号の説明】

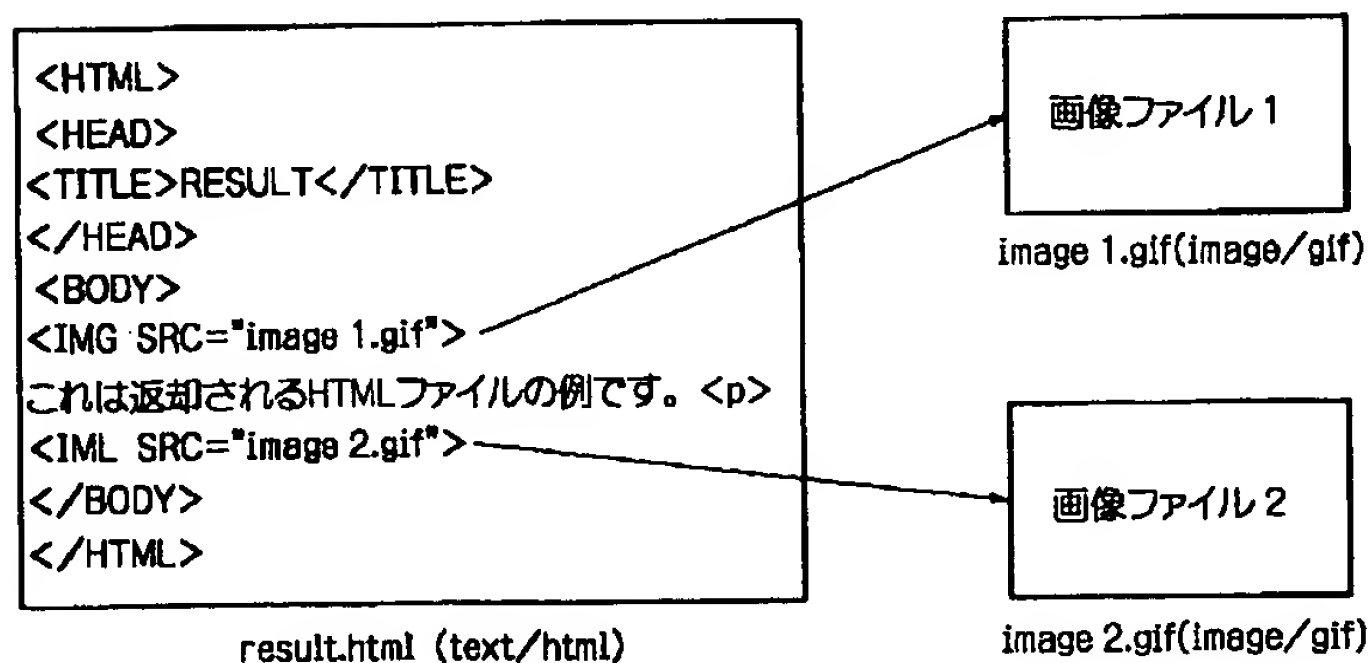
- 1 閲覧装置
- 2 Webサーバ
- 11 Webアクセス部
- 12 結果解析部
- 13 他メディア処理部
- 14 テキスト処理部
- 15 URL管理部
- 16 文字コード識別部

- 14 テキスト処理部
- 15 URL管理部
- 16 文字コード識別部
- 17 ユーザインタフェース部
- 18 識別済みURL DB
- 21 イメージ処理部
- 22 音声処理部
- 23 動画処理部
- 24 他AP処理部
- 10 31～35, 41, 42 ステップ
- 51 入力装置
- 52, 53 記憶装置
- 54 表示装置
- 55 送信装置
- 56 記録媒体
- 57 データ処理装置

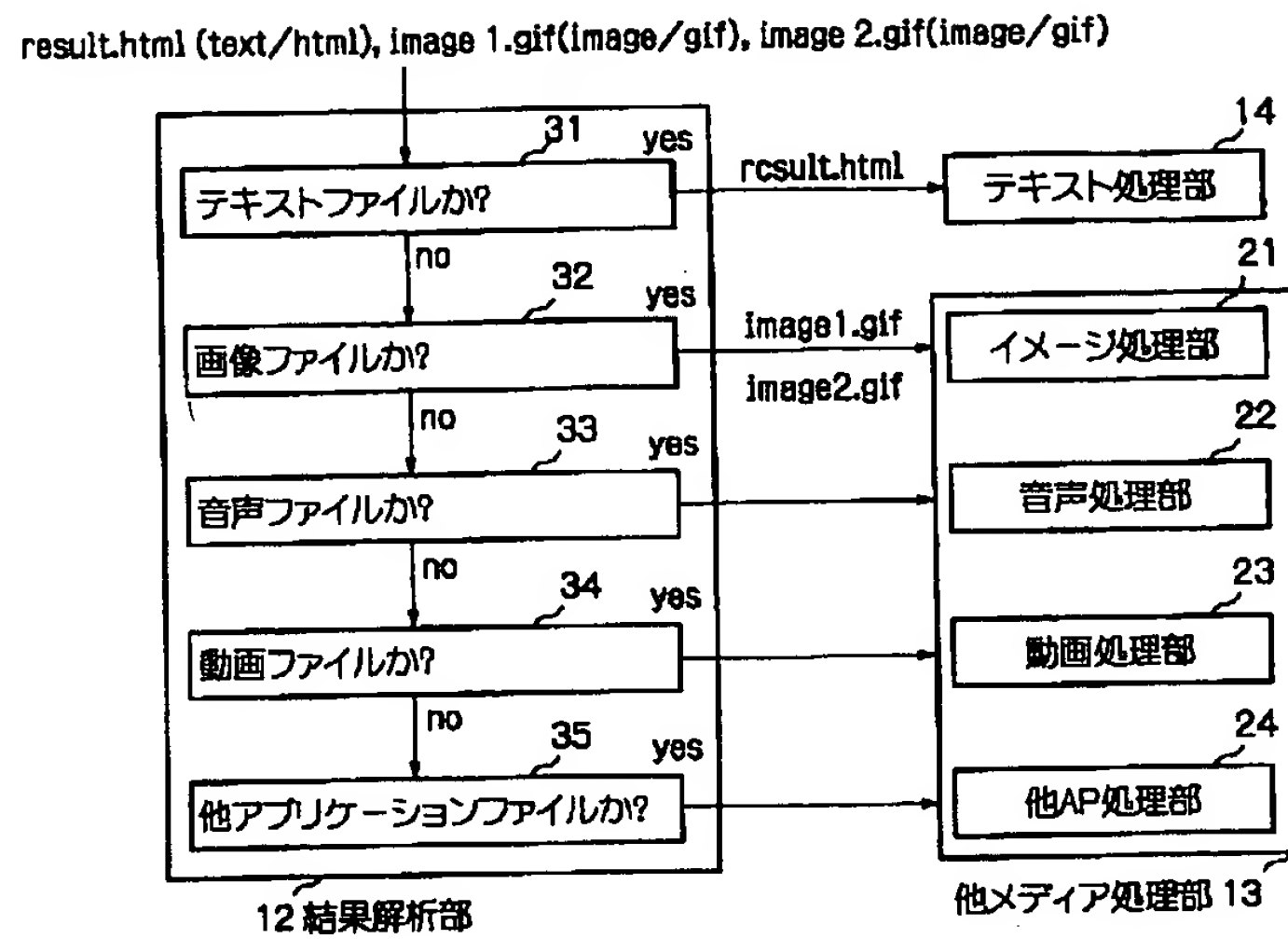
【図1】



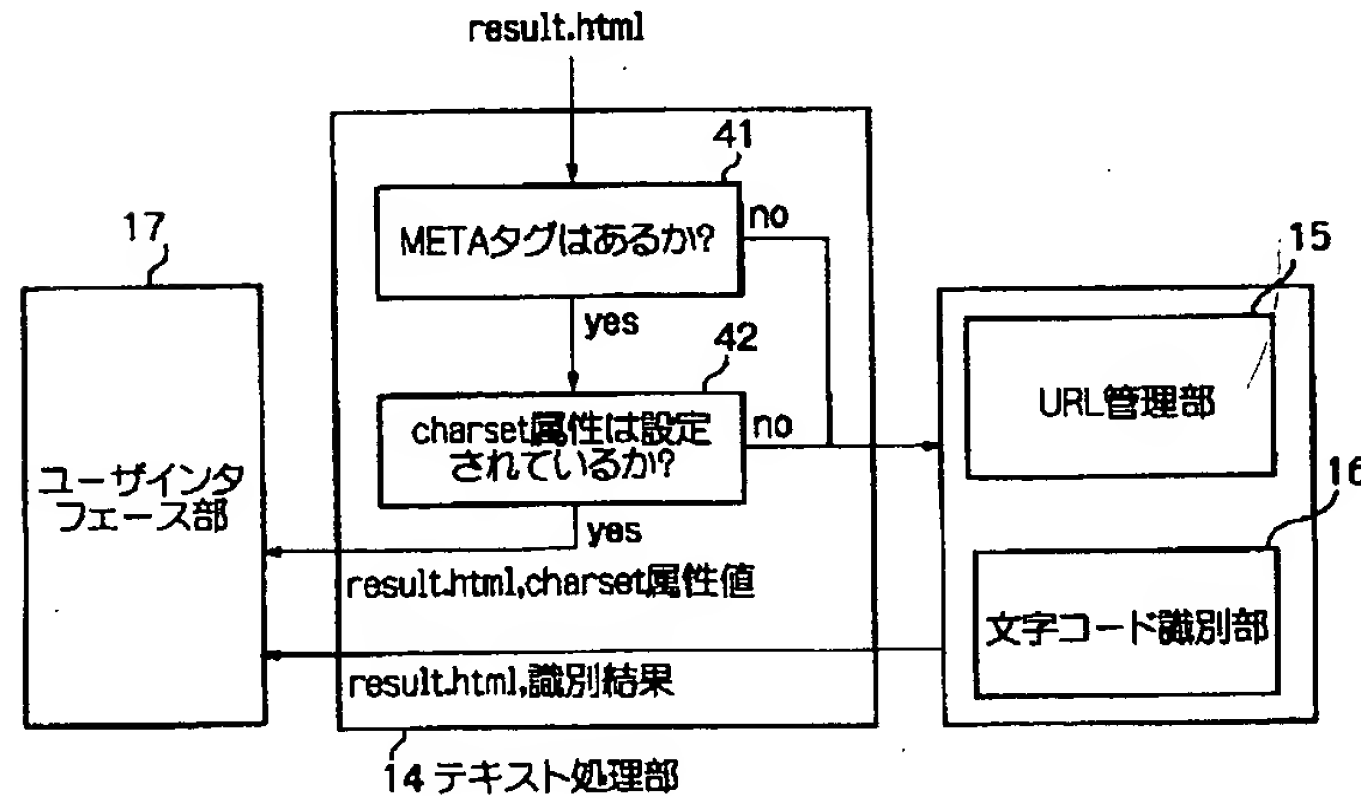
【図2】



【図3】



【図4】



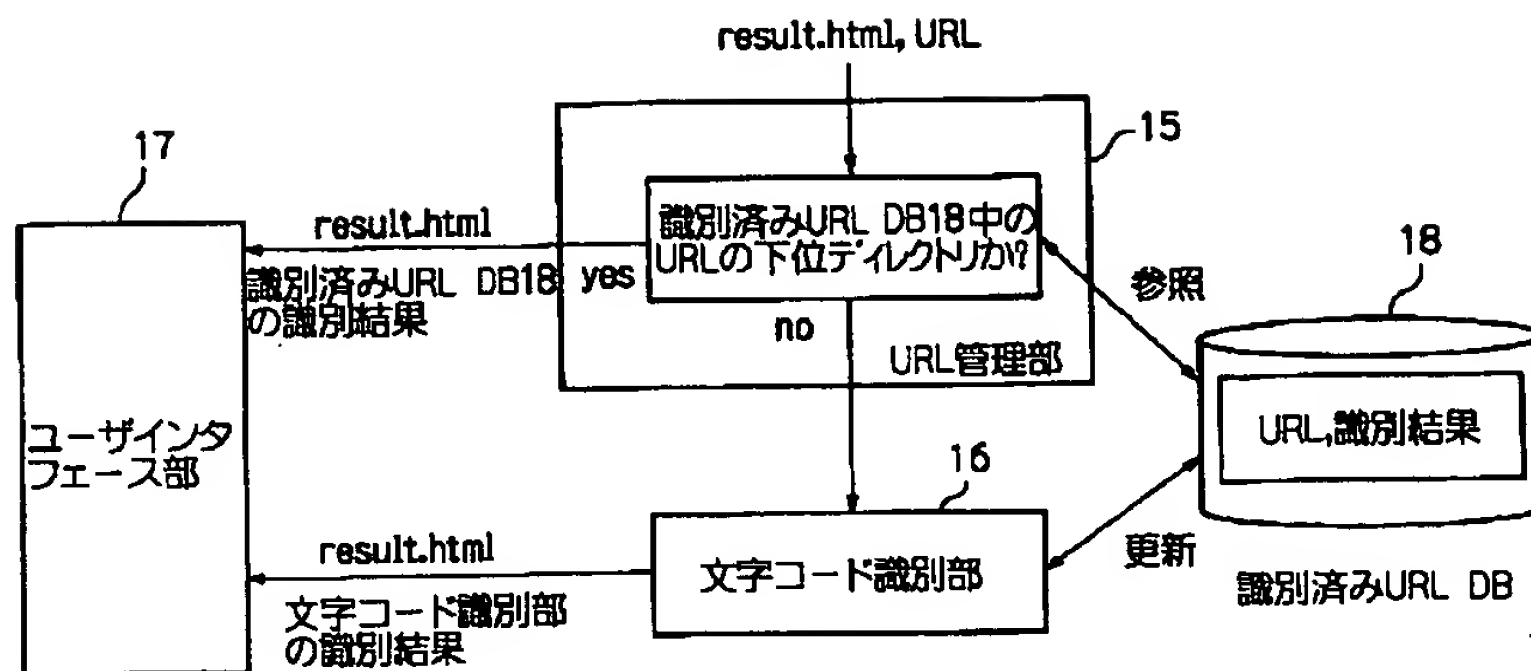
【図5】

```

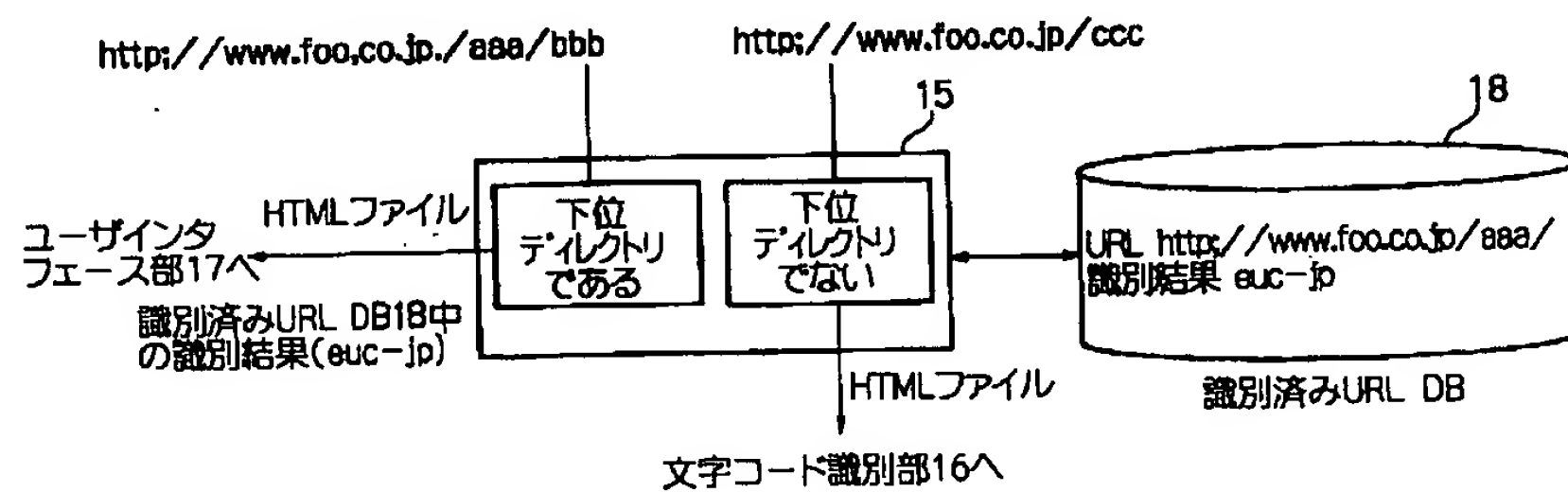
<HTML>
<HEAD>
<TITLE>TEST1</TITLE>
<METAHTTP-EQUIV="Content-Type"CONTENT="text/html; charset=x-euc-jp">
</HEAD>
...
</HTML>
  
```



【図6】



【図7】



【図8】

